Reference priors of nuisance parameters in Bayesian sequential population analysis

Nicolas Bousquet*

October 12, 2010

Abstract

Prior distributions elicited for modelling the natural fluctuations or the uncertainty on parameters of Bayesian fishery population models, can be chosen among a vast range of statistical laws. Since the statistical framework is defined by observational processes, observational parameters enter into the estimation and must be considered random, similarly to parameters or states of interest like population levels or real catches. The former are thus perceived as nuisance parameters whose values are intrinsically linked to the considered experiment, which also require noninformative priors. In fishery research Jeffreys methodology has been presented by Millar (2002) as a practical way to elicit such priors. However they can present wrong properties in multiparameter contexts. Therefore we suggest to use the elicitation method proposed by Berger and Bernardo to avoid paradoxical results raised by Jeffreys priors. These benchmark priors are derived here in the framework of sequential population analysis.

La grande variété des lois *a priori* utilisables pour refléter l'incertitude sur un paramètre, au sein d'un modèle de dynamique de population traité dans un cadre statistique bayésien, nécessite de pouvoir bénéficier d'une distribution de référence, dite neutre (non-informative), à laquelle les comparer, en particulier pour juger l'impact de la prise en compte d'une connaissance parfois subjective. Dans cette optique, Millar (2002) a calculé l'*a priori* de Jeffreys, qui possède la propriété d'invariance par reparamétrisation, pour différents modèles simples de population utilisés en halieutique. Cependant, cet *a priori* présente de mauvaises propriétés lorsque le nombre de paramètres est grand. Dans cette note, nous suggérons d'utiliser l'*a priori* de référence de Berger-Bernardo pour les paramètres de nuisance liés au processus d'observation, dans le cas où l'on modélise les variations de population par des modèles séquentiels de grande dimension.

Introduction

From years Bayesian statistics have been recognized as a practical methodology allowing to take account of natural uncertainties arising in biological population models. They have been applied to a growing number of problems in fishery research (Punt and Hilborn 1997). In most applications a large part of parameters can benefit from expert knowledge, such that prior distributions can be elicited in order to increase the data information summarized in the likelihood of observations. It is essential, however, that prior elicitation must be carefully led about other parameters for which no information is available.

To avoid integrating unduly subjective information into Bayesian inference, the need for noninformative priors in fishery research has been highlighted by Millar (2002). He reviewed formal rules to define and elicit such priors in the Bayesian world and proposed

*EDF Research and Development, 6 quai Watier, 78401 Chatou, France.

using Jeffreys priors. Indeed, unlike the common but tricking flat priors which lead to paradoxes (Kass and Wasserman 1996), Jeffreys priors have the property to be independent on parametrization choices.

In many cases this property appears desirable. For instance, when the scale of an initial population is unknown (since nonobservable) and is considered as a parameter to be estimated, any subjective prior (even diffuse) is expected to influence the posterior estimation not only of the scale but of many other important quantities directly linked to the scale, like the maximum sustainable yield, in ways predictable with difficulty. A scale-invariant prior can solve such issues. Therefore Millar (2002) derived Jeffreys prior for some common surplus production models and biological parameters of sequential population models. Eliciting noninformative priors has another interest: the Bayesian analyst can aim to understand the amount of subjective information used for an assessment, comparing subjective and objective posterior distributions.

If Jeffreys prior is quite useful and practical for a single parameter, it can however suffer from strong deficiencies in multiparameter problems. Cases of posterior inconsistency have been highlighted by Berger and Bernardo (1992). When deriving Jeffreys prior for sequential population models, Millar (2002) warned his readers about this knotty feature. Then reference priors were elicited to preserve the desirable invariance properties of the Jeffreys priors, but to avoid such paradoxical consequences (Koop et al. 2007, chap.8). The term "reference priors" was used by Millar (2002) for referring to Jeffreys prior, but is used here according to the usual terminology of Bayesian statistics, referring to the elicitation method proposed originally by Bernardo (1979) then refined by Berger and Bernardo (1992).

In facts, difficulties mainly occur when parameters are hierarchised. This was formally demonstrated by Ghosal (1997). In fishery research, the main issue is evaluating the population level over the time steps. It requires to focus on real catch values to get absolute evaluations. In a classical statistical frame, population levels and catches can be viewed as parameters to estimate or hidden states of Markov chains whose trajectories must be reconstituted. In the Bayesian framework these differences become blurred and we can simply define them as *interest parameters* and denote their vector as θ_I.

However, some parameters always enter into population models because of observational processes. Since they are unknown, complicate the inference and because we have less interest in estimating them than estimating population levels and catches, they take the sense of *nuisance parameters* (Idier 2001, chap. 3). Typically, they are at least three: the variance ψ^2 arising from the direct observation of catches, the variance ϕ^2 linked to the indirect observation of survey indices and the global catchability q linking the real survey indices and the population levels. Note that other parameters, for instance linked to recruitment or selectivity, notably differ from $\theta_N = (\psi^2, \phi^2, q)$ since they arise from biological mechanisms or technical features that are independent on the observational experiments. As an example, selectivity parameters intrinsically characterize the state of a fishnet, and some external knowledge is often available to provide a prior range of most probable values (Harley and Myers 2001). Thus, θ_N differs from other parameter vectors since it remains relative to a given experiment, and its prior distribution with density $\pi(\theta_N)$ has to be considered noninformative.

Therefore we suggest in this note to adopt the reference prior approach for eliciting benchmark priors for θ_N, which completes the results given by Millar (2002). In the following, a common sequential population model is defined through its observational equation. No hypothesis is made on the linkage between populations and catches (for instance Baranov equations or their Pope's approximation). Then we give the definition of reference priors and provide our elicited prior $\pi(\theta_N)$ for nuisance parameters.

Sequential population model

Sequential population models present numerous variants, for instance depending on how natural mortality parameters are build, assumed to be constant or varying in function of some age classes. However, the following formulation encompasses a variety of modellings (see for instance Chassot et al. 2009), especially cohort models.

Survey information. For $a = 1, \ldots, A$ (age classes) and $t = 1, \ldots, T$ (time steps), let $I_{a,t}$ denote a survey index such that $I_{a,t} = qs_a N_{a,t}$ where s_a denotes the selectivity-by-age characterizing survey gears and $N_{a,t}$ is the abundance of the studied species. Correlated observational errors are assumed between ages (Myers and Cadigan 1995), following classical lognormal distributions: denoting $I^*_{a,t}$ the observed index, one assumes

$$J^*_{a,t} \;=\; \log I^*_{a,t} \;=\; \log I_{a,t} + \epsilon_{a,t} + \eta_t \tag{1}$$

where $\epsilon_{a,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and $\eta_t \overset{iid}{\sim} \mathcal{N}(0, \tau^2)$. The survey likelihood then arises from the T equations

$$J^*_t \;=\; \sum_{a=1}^{A} J^*_{a,t} \;= \tilde{J}_t + \; A \log q + \epsilon'_t, \qquad t = 1, \ldots, T, \tag{2}$$

where $\tilde{J}_t = \sum_{a=1}^{A} \log s_a N_{a,t}$, $\epsilon'_t \overset{iid}{\sim} \mathcal{N}(0, \phi^2)$ and $\phi^2 = A\sigma^2 + A^2\eta^2$.

Catch information. Following Aanes et al. (2007), we adopt the following formulation for the variations of observed total catches C^*_t over ages. Denote C_t the real total catch at time t. Then the catch likelihood arises from the T equations

$$C^*_t \;=\; C_t \exp(\nu_t), \qquad t = 1, \ldots, T, \tag{3}$$

where $\nu_t \overset{iid}{\sim} \mathcal{N}(-\psi^2/2, \psi^2)$. The lognormal error ν_t has mean $-\psi^2/2$ and not 0 to ensure $\mathrm{E}[C^*_t] = C_t$. Otherwise, estimation naturally underestimates the real catches, making worse the frequent bias due to underreported landings C^*_t (Bousquet et al. 2010). Such a correction (usually called Laurent correction) is not useful in the survey equation (1) since the bias $\mathrm{E}[I^*_{a,t}] - I_{a,t}$ is only reflected in the estimation of q (Aanes et al. 2007).

A condensed parametrization of the full observation model is given by (θ_I, θ_N) where $\theta_I = (\tilde{J}_1, \ldots, \tilde{J}_T, \log C_1, \ldots, \log C_T)$. To be very general, one assumes that an informative prior measure $\pi(\theta_I|\theta_N)$ can always be elicited (possibly independent on θ_N). As proposed by Follestad (2003), this prior can encompass environmental fluctuations $\mu^{(i)}_{a,t}$, $i = 1, 2$ (commonly gaussian), such that

$$\begin{aligned} C_{a,t} &= f^{(1)}(N_{a,t}, F_{a,t}) \exp(\mu^{(1)}_{a,t}), \\ N_{a+1,t+1} &= f^{(2)}(N_{a,t}, F_{a,t}, M_{a,t}) \exp(\mu^{(2)}_{a,t}) \end{aligned}$$

where the $f^{(i)}$ are deterministic functions (such as well-known Baranov or Pope's equations when $A > 1$, or simpler surplus-production relations if a unique age class is considered) and (F, M) denote fishing and natural mortality rates, and external information on biological parametric functionals (like (F, M)).

Eliciting the nuisance prior

We are here interested in eliciting the Berger-Bernardo reference prior $\pi(\theta_N)$, independently from any prior choice on θ_I. Reference priors differ from Jeffreys priors in the sense they are defined as the optimal priors according to an information-theoretical criterion, while Jeffreys priors emerge from a formal definition of a so-called ideal feature (invariance to reparametrization).

To be clearer, denote $f(\mathbf{x}|\theta_N)$ the density of observed data $\mathbf{x}$ (here the (C_t^*, J_t^*)), unconditionally to θ_I, and let θ_N takes its values in the metric space Θ. The reference prior is the unique maximizer of the Kullback-Leibler divergence between the posterior and the prior densities

$$\mathrm{KL}(\pi|\mathbf{x}) \quad = \quad \int_{\Theta_N} \pi(\theta_N|\mathbf{x}) \log \frac{\pi(\theta_N|\mathbf{x})}{\pi(\theta_N)} \, d\theta_N$$

on average on data that are *marginally* sampled, i.e. $\mathbf{x} \sim m(\mathbf{x}) = \int_{\Theta_N} f(\mathbf{x}|\theta_N)\pi(\theta_N)\, d\theta$, and assuming finally asymptotic conditions. In summary the reference prior is defined as

$$\pi^* \quad = \quad \arg\max_{\pi} \left\{ \lim_{\text{card}(\mathbf{x})\to\infty} \mathrm{E}_m \left[\mathrm{KL}(\pi|\mathbf{X})\right] \right\} \tag{4}$$

The Kullback-Leibler distance between posterior and prior measures the (relative) Shannon's information quantity that is yielded by the data into the inference (independently of any parametrization); maximizing it asymptotically is similar to minimizing the quantity of information yielded by the prior independently on the number of data. Hence the prior weight is minimized on average into the posterior. Further details can be found in Cover and Thomas (2006).

Note that asymptotism makes sense in our framework, since the dimension of θ_N does not increase with the number $2T$ of observed data. One would have to be careful in the elicitation of a conditional reference prior $\pi(\theta_I|\theta_N)$ since the dimension of θ_I always linearly increases with T, because we consider the catches as parameters (besides this invalidates any frequentist estimation theory over θ_I, see Bousquet et al. 2010). Some relations (like $F-$constraints, cf. Gavaris and Ianelli 2002) can often be established to diminish the dimension of θ_I. But anyway our final result $\pi(\theta_N)$ will not depend on any hypothesis placed on the intern covariance of θ_I. Finally, we give the reference prior in next theorem. Its proof is given in Appendix.

Theorem 1. The solution of the optimization problem (4) is

$$\pi^*(\psi^2, \phi^2, q) \quad \propto \quad \psi^{-3}\phi^{-3}q^{-1} \, \mathbb{1}_{\{(\psi,\phi,q)\in \mathbb{R}^3_{+,*}\}}.$$

The elicited prior is clearly improper except if some constant values other than 0 and ∞ can bound the parameters. Note that $\pi^*(\psi^2)$ and $\pi^*(\phi^2)$ differ significantly from the familiar Jeffreys priors that can be derived from Millar (2002):

$$\pi(\psi^2) \quad \propto \quad \psi^{-2}, \quad \pi(\phi^2) \; \propto \; \phi^{-2}.$$

Giving more importance to smaller values of observational variances, they transmit more information to the estimation of parameters of interest. This is not the case for $\pi^*(q)$ which is found similar to the classic scale-invariant measure (Millar 2002).

Remark. Such priors remain conjugate conditionally to the set of all random quantities and observations, generically denoted $\mathcal{D}$. Indeed one has

$$\begin{aligned}
\log q|\psi^2, \mathcal{D} &\sim \mathcal{N}\left(\frac{1}{AT}\sum_{t=1}^{T}\sum_{a=1}^{A}\log\frac{I^*_{a,t}}{s_a N_{a,t}}, \frac{\psi^2}{A^2 T}\right), \\
\psi^2|q, \mathcal{D} &\sim \mathcal{IG}\left(\frac{T+1}{2}, \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{a=1}^{A}\log\frac{I^*_{a,t}}{q s_a N_{a,t}}\right]^2\right), \\
\sigma^2|\mathcal{D} &\sim \mathcal{IG}\left(\frac{T+1}{2}, \frac{1}{2}\sum_{t=1}^{T}\left[C^*_t - C_t\right]^2\right),
\end{aligned}$$

which favors using within-Gibbs algorithms in any posterior computation of quantity of interest.

Conclusion

In this note we offer a counterpart to the choice of parametrization-invariant Jeffreys priors to take account of the variability of nuisance parameters (observational variances and global catchability) in sequential population models. Our prior is elicited to avoid potential incoherences arising from Jeffreys prior in high-dimensional models, while other priors elicited using past knowledge (or more generally *expert* knowledge) are placed on parameters of interest. As a new entrance into the catalog of reference priors for fisheries models, this note fits in line with the wishes expressed by Millar (2002) in his seminal work.

References

[1] Aanes, S., Engen, S. Saether, B.-E., Aanes,R. (2007). Estimation of the parameters of fish stock dynamics from catch-at-age data and indices of abundance: can natural and fishing mortality be separated? *Can. J. Fish. Aquat. Sci.*, 64: 1130-1142.

[2] Berger, J.O., Bernardo, J.M. (1992). On the development of reference priors (with discussion). In: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Eds., *Bayesian Statistics 4*, Oxford University Press, pp. 35-60.

[3] Bernardo, J.M. (1979). Reference Posterior Distributions for Bayesian Inference, *J. R. Statist. Soc.*, **41**, pp. 113-147.

[4] Bousquet, N., Duchesne, T., Rivest, L.-P. (2008). Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise, *Journal of Theoretical Biology*, 254(1): 65-75.

[5] Bousquet, N., Duchesne, T., Cadigan, N., Rivest, L.-P. (2010). Detecting and correcting underreported catches in fish stock assessment: trial of a new method. *Can. J. Fish. Aquat. Sc.*, 67(8): 1247-1261.

[6] Chassot, E., Duplisea, D., Hammill, M., Caskenette, A., Bousquet, N., Lambert, Y., Stenson, G. (2009). Role of predation by harp seals *Pagoliphus groenlandicus* in the collapse and non-recovery of northen Gulf of St. Lawrence cod *Gadus morhua*. *Mar. Ecol. Prog. Series*, 379: 279-297.

[7] Cover, T.M., Thomas, J.A. (2006). *Elements of Information Theory* (2nd edition). Wiley.

[8] Follestad, T. (2003). Stochastic modelling and simulation based inference of fish population dynamics and spatial variation in disease risk. *Dr Ing. Thesis*. Norwegian University of Science and Technology.

[9] Gavaris, S., Ianelli, J.N. (2002). Statistical Issues in Fisheries' stock assessment. *Scandinavian Journal of Statistics*, 29: 245-267.

[10] Ghosal, S. (1997). Reference priors in multiparameter nonregular cases. *Test*, 6:159-186.

[11] Ghosh, J.K., Mukerjee, R. (1992). Noninformative priors (with Discussion). In: *Bayesian statistics 4*, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith, Oxford University Press.

[12] Harley, S.J., Myers, R.A. (2001). Hierarchical Bayesian models of length-specific catchability of research trawl surveys, *Can. J. Fish. Aquat. Sci.*, 58: 1569-1584.

[13] Idier, J. (*under the direction of*) (2001). Approche bayésienne pour les problèmes inverses (in French). Lavoisier.

[14] Kass, R.E., Wasserman, L. (1996). The selection of prior distributions by formal rules, *Journal of the American Statistical Association*, **91**, pp. 1343-1370.

[15] Koop, G., Poirier, D.J. Tobias, J.L. (2007). *Bayesian Econometric Methods*. Cambrigde University Press.

[16] Millar, R.B. (2002). Reference priors for Bayesian fisheries models, *Can. J. Fish. Aquat. Sc.*, 59:1492-1502.

[17] Myers, R.A., Cadigan, N.G. (1995). Statistical analysis of catch-at-age data with correlated errors. *Can. J. Fish. Aquat. Sci.*, 52:1265-1273.

[18] Punt, A.E., Hilborn, R. (1997). Fisheries stock assessment and decision analysis: the Bayesian approach. *Rev. Fish. Biol. Fish.*, 7: 35-63.

[19] Sun, D., and Berger, J.O. (1998). Reference priors with partial information, *Biometrika*, 85: 55-71.

Appendix: proof of Theorem 1

We start from an asymptotic development of $\mathrm{E}_m[\mathrm{KL}(\pi|\mathbf{X})]$, under classic regularity conditions, provided by Ghosh and Mukerjee (1992) and recalled by Sun and Berger (1998, Equ. 2). Recalling that the size of our observed data is $2T$, we have (up to an additive constant)

$$\mathrm{E}_m[\mathrm{KL}(\pi|\mathbf{X})] \quad = \quad \frac{3}{2}\log T - \int_{\Theta_N} \pi(\theta_N)\log\frac{\pi(\theta_N)}{\pi^J(\theta_N)}\,d\theta_N + o(1) \tag{5}$$

where

$$\pi^J(\theta_N) \quad \propto \quad \exp\left\{\frac{1}{2}\int_{\Theta_I} \pi(\theta_I|\theta_N)\log\frac{|\Sigma|}{|\Sigma_2|}\,d\theta_I\right\}, \tag{6}$$

(Θ_I being the parameter space for θ_I), $|\Sigma|$ is the determinant of the per observation Fisher information matrix

$$\Sigma(\theta) \quad = \quad \mathrm{E}_X\left[\frac{\partial^2 \log f(X|\theta)}{\partial\theta\partial\theta'}\right]$$

and $\Sigma_2(\theta_I|\theta_N)$ is this same matrix for θ_I, given that θ_N is held fixed. One has (X^* being a generic notation for (C_t^*, J_t^*))

$$\begin{aligned}
\frac{\partial^2 \log f(X^*|\theta)}{\partial \log^2 C_t} &= -\frac{1}{\psi^2}, \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial \log C_t \partial \psi^2} &= -\frac{1}{\psi^4}(\log C_t^* - \log C_t), \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial (\psi^2)^2} &= \frac{1}{2\psi^4} - \frac{(\log C_t^* - \log C_t)^2}{\psi^6}, \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial (\phi^2)^2} &= \frac{1}{2\phi^4} - \frac{1}{\phi^6}\left(J_t^* - A\log q - \tilde{J}_t\right)^2, \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial \log q \partial \phi^2} &= -\frac{A}{\phi^4}\left(J_t^* - A\log q - \tilde{J}_t\right), \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial \phi^2 \partial \tilde{J}_t} &= -\frac{1}{\phi^4}\left(J_t^* - A\log q - \tilde{J}_t\right), \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial (\log q)^2} &= -\frac{A^2}{\phi^2}, \\
\frac{\partial^2 \log f(X^*|\theta)}{\partial \log q \partial \tilde{J}_t} &= -\frac{A}{\phi^2},
\end{aligned}$$

$$\frac{\partial^2 \log f(X^*|\theta)}{\partial \tilde{J}_t^2} = -\frac{1}{\phi^2}.$$

Any other cross derivative is zero. With

$$\begin{aligned}
\mathrm{E}_X\left[\log C_t^* - \log C_t\right] &= -\psi^2/2, \\
\mathrm{E}_X\left[(\log C_t^* - \log C_t)^2\right] &= \psi^2 + \psi^4/4, \\
\mathrm{E}_X\left[\left(J_t^* - A\log q - \tilde{J}_t\right)\right] &= 0, \\
\mathrm{E}_X\left[\left(J_t^* - A\log q - \tilde{J}_t\right)^2\right] &= \phi^2,
\end{aligned}$$

one can write

$$\Sigma = \begin{pmatrix} \phi^{-4}/2 & & (0) \\ & A\phi^{-2}\Sigma_a & \\ (0) & & \psi^{-2}\Sigma_b/2 \end{pmatrix}$$

with

$$\Sigma_a = \begin{pmatrix} A & 1 & 1 & \dots & 1 \\ 1 & 1 & & & \\ 1 & & 1 & & (0) \\ \vdots & (0) & & \ddots & \\ 1 & & & & 1 \end{pmatrix}$$

with dimension $T+1$ and

$$\Sigma_b = \begin{pmatrix} \psi^{-2}(2+\psi^2)/2 & -1 & -1 & \dots & -1 \\ -1 & 2 & & & \\ -1 & & 2 & & (0) \\ \vdots & (0) & & \ddots & \\ -1 & & & & 2 \end{pmatrix}$$

with dimension $T+1$. Note that $\Sigma_b = Q_b^T Q_b$ with Q_b the lower triangular matrix

$$Q_b \quad = \quad \begin{pmatrix} \sqrt{|\psi^{-2}(2+\psi^2)/2 - T/2|} & & & \\ -1/\sqrt{2} & \sqrt{2} & & (\mathbf{0}) \\ -1/\sqrt{2} & & \sqrt{2} & \\ \vdots & (0) & & \ddots \\ -1/\sqrt{2} & & & & \sqrt{2} \end{pmatrix}$$

Thus, we have

$$|\Sigma| \quad \propto \quad \phi^{-(6+2T)} \psi^{-2(T+2)} \left|(T-1)\psi^2 - 2\right|.$$

Finally, note that

$$\Sigma_2 \quad = \quad \begin{pmatrix} \phi^{-2} I_T & 0 \\ 0 & \psi^{-2} I_T \end{pmatrix}$$

where I_T is the unit diagonal matrix of dimension T. Thus

$$|\Sigma_2| \quad \propto \quad \phi^{-(2T)} \psi^{-2T}.$$

Since $|\Sigma|/|\Sigma_2|$ does not depend on θ_2, from (6) we have

$$\begin{aligned} \pi^J\left(\psi^2, \phi^2, \log q\right) \quad &\propto \quad \sqrt{|\Sigma|/|\Sigma_2|}, \\ &\propto \quad \phi^{-3}\psi^{-2}\sqrt{|(T-1)\psi^2 - 2|}. \end{aligned}$$

which can be rewritten (assuming of course $T > 1$)

$$\pi^J\left(\psi^2, \phi^2, q\right) \quad \propto \quad \pi^*\left(\psi^2, \phi^2, q\right)(T-1)\sqrt{\left|1 - \frac{2}{\psi^2(T-1)}\right|}.$$

Denoting $\mathrm{K}_L(\pi_1||\pi_2)$ the Kullback-Leibler divergence between densities π_1 and π_2, the asymptotic criterion (5) becomes (up to an additive constant)

$$\begin{aligned} \mathrm{E}_m[\mathrm{KL}(\pi|\mathbf{X})] \quad &= \quad \frac{3}{2}\log T + \log(T-1) - \mathrm{K}_L\left(\pi||\pi^*\right) + \frac{1}{2}\int_{\Theta_N} \pi(\theta_N)\log\left|1 - \frac{2}{\psi^2(T-1)}\right| \, d\theta_N + o(1), \\ &\sim \quad \frac{5}{2}\log T - \mathrm{K}_L\left(\pi||\pi^*\right) + \frac{1}{2}\int_{\Theta_N} \pi(\theta_N)\left(1 + O(T^{-1})\right) \, d\theta_N \end{aligned}$$

when T increases. Since the Kullback-Leibler divergence is always positive or null, one finally must choose $\pi = \pi^*$ to maximize this asymptotic expression. Indeed, $\mathrm{K}_L(\pi||\pi^*) = 0$ if and only if $\pi = \pi^*$ (cf. Cover and Thomas 2006).